
The NIST Open MT team concluded the normal evaluation cycle for MetricsMATR 2008 with a workshop on October 25th. This document outlines our current plans for ongoing analysis.

Date: November 12, 2008

(0) Availability of workshop talks and system descriptions

NIST will be asking all presenters to provide NIST with an electronic copy of their presentation. We will create PDF versions of the talks and distribute them to all workshop participants and all MetricsMATR participants.

NIST will distribute the system descriptions to all workshop participants and all MetricsMATR participants. AMTA has requested these descriptions, and we have obliged.

(1) Updated metric submissions

NIST has received updated metric submissions and we have two metrics that we were unable to install during the official evaluation period.

We will not be running additional metrics over the evaluation data set, but will include them (when appropriate) as runs over the development set and other publically available data.

(2) Scoring the MetricsMATR development data set

NIST is in the process of scoring the MetricsMATR development data with each metric. NIST will distribute the full scores reports to the evaluation participants. This will serve two purposes; first – a sanity check for each developer that we are running each metric as intended; second – developers will have data and scores for analysis and can determine which metrics were tuned on the development data.

Anticipated date of release: no later than Friday November 21st, 2008

(3) Releasing MetricsMATR scoring code

NIST is in the process of fine-tuning our report generation software, adding some missing functionality. We intend to distribute a package that is capable of generating the same type of reports. Developers may be able to use such a package to compare multiple versions of their metrics with selected baseline metrics.

Anticipated date of release: no later than Friday November 21st, 2008

(4) Online release of results

The online report is currently hidden (no links to it). NIST will soon link to the results from the MetricsMATR web site. The results will reflect the metrics as submitted and the human assessments as they existed at the time of the workshop. Further annotation of the references may/or may not be reflected in future additive updates to the online results.

Some planned additive updates will include confidence intervals, statistical significance sensitivity, and genre color-coded scatter plots.

(5) Releasing anonymously MetricsMATR's evaluation data with metric scores

IBM has suggested that NIST could release for every segment (system and human translations) all human judgments and all submitted automatic scores, without release the actual sentences (words).

We prefer an alternative (#6) below.

(6) Releasing WMT08 data scored with MetricsMATR metrics

JHU suggested that we run the MetricsMATR metrics over the WMT08 evaluation data and release the scores and their human judgments.

We are in the process of generating these score files. They will be released via the MetricsMATR web site.

(7) Metric to metric correlations

NIST will produce a matrix of metric-to-metric correlations. Metrics that correlate well to adequacy but not to each other may reveal opportunities for metric improvement.

Internally NIST will continue analysis. We will investigate alternative methods of aggregating segment assessments to document and system scores, and alternative methods of processing the evaluation data with each metric. We will continue to look for metric anomalies and report back to the developers. The human assessments will receive a closer look in order to expose judgments that may be an obvious error.